基于语言结构功能的音位负担计算方法*

孔江平[1]　李永宏[2]

1. 北京大学中文系中国语言学研究中心　北京　jpkong@pku.edu.cn
2. 西北民族大学中国民族信息技术研究院

提要　根据结构语言学的音位对立理论和以各种语言语料库为基础，进行了一系列的理论研究和数据分析，提出了音位结构功能负担的计算方法。计算步骤主要包括：1）音节对立频次统计；2）对立类型负担量计算；3）音位系统的负担量计算；4）音位的平均负担量计算。利用本方法，以汉语 20 个方言 3000 音节为研究对象，计算了对立类型负担量、音位系统负担量和音位的平均负担量在方言上的分布，并提出了音位负担量计算的基本理论。本方法可应用于语言音位功能研究、语言历史音变和演化研究以及语音工程的应用研究。

关键词　音位　功能负担　音位负担　音位结构

零　引言

音位（phoneme）是结构主义音位学（phonemics）的一个概念，是指语言中区别意义的最小语音单位。建立一种语言的音位系统，主要基于对立原则、互补原则、相似性原则、经济性原则等等（布龙菲尔德 1980，萨皮尔 1985，布洛赫、特雷杰 1965）。结构主义的音位学在语言的基本描写和演化的研究方面起了重要的作用，是现代语言学的基础。然而，由于不同语言音位系统的性质的差异，结构主义的音位学在音位功能的解释方面显然存在许多不足的地方。例如，汉藏语系语言主要是单音节结构，由声韵调组成，而印欧语系语言主要是多音节结构，由音素构成。另外，一种语言中有的音位只出现在极个别词中，而且系统性很差。由此看出，结构主义音位的定义对音位在一个语言中的频度和功能都无法很好的解释。这直接导致了利用目前的音位学理论很难解释许多语言信息传递的问题、语言演化的问题和语言类型的问题。另外，在音位的描写和分类方面，有些学者从生理发音的角度来描写，有些学者注重声学的作用，用声学区别性特征来描写，还有些学者强调心理的功能，利用感知结果来描写。因此，在音位学研究中存在种种不同的观点和争论。为此赵元任先生专门讨论了“音位标音法的多能性”（赵元任 1934/1985）来协调不同的争论。

音位学理论上的这些问题实际上一直受到语言学家的关注，其中音位功能负担（functional load）的研究就是试图解决这些问题的一个方面。音位负担量的概念和研究可以追溯到早期的布拉格学派时期（Mathesius 1929，Jakobson 1931，Trubetzkoy 1939)，当时主要注重于音位学的二元对立。在功能负担的语言学研究方面，50 年代，主要有霍凯特（Hockett 1955，1967）和格林博格（Greenberg 1959）的研究。霍凯特认为：功能负担的重要性在于它对描写音韵系统有重要的价值，从而使我们可以有一个尺度来认识语言信息、语言冗余度和言语识别。格林博格认为：功能负担以通用的方式反映了一组音位或一组对立特征各成员

* 本研究得到国家社科重大基金项目（项目编号：10&ZD125）支持。

之间对有区别意义信号的贡献。在 60 年代，主要有赫厄希斯瓦尔德（Hoenigswald 1960）关于功能负担和音变的研究，他认为：功能负担和语言的音变有关，并提出了一个假说，即“在一种语言里，如果一种对立用的很少，它的消失对系统造成的危害要小于功能负担大的对立”。60 年代还有王士元教授（Wang 1967）有关功能负担的著名研究和京•罗伯特（King 1965，1967a，1967b）的研究。王士元教授首次实现了功能负担的计算，指出了计量功能负担的困难，并给出了解决这些困难的方法。他讨论了音位系统中常见的三种分布、霍凯特与格林博格的测量方法以及这些方法和香农等（Shannon and Weaver 1949；Shannon 1951；Kucera 1963）的通信理论及各种语言学概念的关系，系统地发展了四种计量功能负担的方法。王士元教授的研究为后来功能负担的研究建立了一个理论上的基本框架。另外，王士元教授还指出“如果功能负担在音变中确实起作用的话，那么用量化的解释至少可以从一个方面阐明音变这一难题”。京• 罗伯特将音变和功能负担一同进行研究，并着重研究了音位功能和语音音变的关系。在本世纪初有苏仁德兰和尼育基（Surendran and Niyogi 2003)、苏仁德兰和利佛（Surendran and Levow 2004)的研究，苏仁德兰和利佛在其研究中不仅讨论了霍凯特的定义，还讨论了音位、区别性特征和超音段特征的功能负担，同时，他们还研究了汉语声调的功能负担，发现汉语声调的功能负担与元音同样高。

音位功能负担的研究促进了语言功能的研究，同时也促进了计算语言学和语音技术的发展。例如，目前语音识别和合成中常用的双音子（diphone）和三音子（triphone）就来源于音位功能负担的理论。从以往的研究看，功能负担研究都是在大文本的基础上进行音位的功能负担统计和量化以及计算该语言的熵值和冗余度。众所周知，世界上大部分语言是没有文字和文献的，这就大大限制了这一理论方法在语言学研究中的应用和发展。因此发展出一种只利用基本语素和词汇就能对语言的音位功能进行计算和量化研究的方法就显得十分迫切。根据结构语言学的音位对立理论和基于我们建设的汉语普通话、汉语方言、藏语方言、藏缅语等语料库，我们进行了一系列的理论研究和数据分析，最终提出了音位结构功能负担的计算方法。这种音位内部的“结构功能负担量”是封闭性的。从性质上看，结构功能负担更能体现一个语言音位系统的性质。因为只需要语言的基本语素便可进行研究，所以能应用于任何一种语言的音位功能研究、语言历史音变和演化研究以及语音合成和识别的研究，并希望能建立和形成一种新的音位功能研究的理论框架。本文以汉语方言为例重点介绍音位结构功能负担的具体计算方法。

壹　音位负担的定义和类型

在音位负担的定义上，我们将两个单音节语素之间的区分定义为语言的一个负担量单位，由一个或多个不同的音位来负担。通常大多数语言用音素音位描写，但中国汉藏语系语言在音位结构上有声母、韵母和声调这一层级的音位，音位单位可以是音素音位或声韵调音位。为了能更好地解释音位负担量的计算方法和汉藏语系语言的音位负担，我们选用声、韵和调作为音位负担量的计算单位。如果由一个音位单位来区别这两个语素，该单位就承载这个语素的音位负担量；如果是由多个音位单位来区别这两个语素，音位负担量由多个音位共同承担。

1.1 定义

一种语言音位系统内部音位结构、音位分布和音位功能的负担，称为“音位结构功能负担”（phoneme structural functional load），简称“音位负担”（phoneme load）。对音位结构功

能负担的相关研究称为“音位负担研究”，主要是以一种语言基本语素的内部音位结构、音位分布和音位功能负担为研究对象。“言语功能负担研究”（functional load）是以一种语言的实际文本为基础的音位功能负担研究，两者有本质的区分

为了能对计算方法进行清楚地说明，我们根据汉藏语系语言单音节的特点，提出了一些新的概念并进行详细解释。

语素:本文是指单音节语素。

音节对立：是指不同单音节语素的对立，可体现为某两个单音节语素的声母对立、韵母对立和声调对立以及这些对立的组合。

语言负担量单位：一个单音节语素承载 1 个语言负担量单位。

音节负担量：在一个封闭的音位系统和基本语素系统内，一个音节的负担量为该音节负担量的总和，即该音节同音字的数量。

1.2 类型

在汉藏语系语言中，基本语素大多是单音节，每个音节均由声母和韵母或声母、韵母和声调组成。音节和音节之间产生 4 种对立类型、8 种对立方式，其中包括：1 种三项对立、3 种两项对立、3 种单项对立和 1 种无对立，形成了音位系统结构和分布的基本形式和框架，见表 1。

表 1 音位对立类型表

对立类型	**对立方式**
三项对立	声/韵/调对立
两项对立	声/韵对立、声/调对立、韵/调对立
单项对立	声母对立、韵母对立、声调对立
无对立	声/韵/调相同

我们以北京话为例来解释这 8 种对立方式。三项对立是指两个单音节语素之间声韵调都不同，如：八 pa55、笛 ti35。两项对立是指两个单音节语素之间只有两个音位单位不同，即声/韵不同，如：八 pa55，低 ti55 或声/调不同，如：八 pa55，大 ta51；或韵/调不同，如：八 pa55，不 bu51。单项对立是指两个单音节语素之间只有一个音位单位不同，即声母不同，如：八 pa55，搭 ta55；或韵母不同，如：八 pa55，波 bo55；或声调不同，如：八 pa55，拔 pa35。无对立是指两个单音节语素之间没有音位单位不同，即同音词，如：八 pa55，巴 pa55。根据以上的基本定义，我们在下一节讨论音位负担的计算方法和整个语言学框架。

贰　计算方法

音位的结构负担计算要靠音节和音节之间的对立关系来体现，包括：音节对立频次统计、对立类型负担量计算和音位负担计算。音节对立频次的统计是对立类型负担量计算和音位负担量计算的基础。

2.1 音节对立频次计算

计算每一个语素对应的音节与其他所有语素对应的音节之间的对立关系，即对立频次。下面以汉语方音字汇 3000 字北京话为例进行说明，计算每一个语素的音节对立情况，图 1 为语素“巴(pa55)”对立频次计算图。

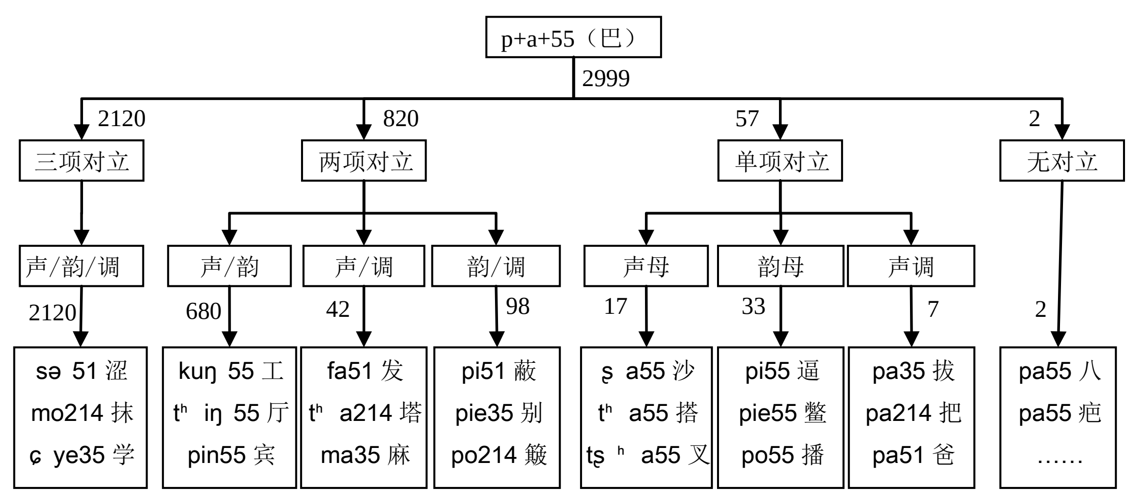


图 1　北京话“巴/pa55/”对立频次计算图

从图 1 中可以看出，“巴”在 3000 个语素中与其他语素共产生 2999 次对立关系，其中 2997 次对立形成不同音节，2 次无对立形成“巴”的同音字，具体如下：

（1）三项对立：声/韵/调对立 2120 次，与“巴”声母、韵母和声调都不相同的语素共有 2120 个。

（2）两项对立：声/韵对立 680 次，即与“巴”声母、韵母不同而声调相同的语素共有 680 个；声/调对立 42 次，即与“巴”声母、声调不同而韵母相同的语素共有 42 个；韵/调对立 98 次，即与“巴”韵母、声调不同而声母相同的语素共有 98 个。

（3）单项对立：声母对立 17 次，即与“巴”声母不同而韵母和声调相同的语素共有 17 个；韵母对立 33 次，即与“巴”韵母不同而声母和声调相同的音节共有 33 个；声调对立 7 次，即与“巴”声调不同而声母和韵母相同的音节共 7 个。

第 1 个音节“巴”计算完成后，用同样的方法计算第 2 个音节，第 3 个音节，以此类推，直到全部 3000 音节计算完毕，得到音节对立频次表。表 2 为北京话 3000 字汇的音节对立频次表。

表 2　北京话音节对立频次表

序号	字目	音节	三项对立	两项对立			单项对立			无对立
			声/韵/调	声/韵	声/调	韵/调	声母	韵母	声调	
1	巴	pa55	2120	680	42	98	17	33	7	3
2	疤	pa55	2120	680	42	98	17	33	7	3
3	八	pa55	2120	680	42	98	17	33	7	3
4	拔	pa35	2112	688	41	121	18	10	9	1
5	把把握	pa214	2320	480	52	103	7	28	9	1
…	…	…	…	…	…	…	…	…	…	…
3000	用	iuŋ51	1714	909	9	236	0	124	7	1
合计			6024398	2151884	226456	337300	82166	137952	25056	14788
平均			2008.1	717.3	75.5	112.4	27.4	46	8.4	4.9

2.2 对立类型负担量计算

对立类型的负担量计算包括三项对立、两项对立和单项对立类型的计算，其大小表示一种语言音节之间的关系及音位组合结构方式。计算过程为：(1)以音节为计算单位，合并同

音字，每个音节的负担量为其同音字个数 T(i)；(2)由于同音字对应的各对立频次数量相同，乘以 T(i)系数，得到音节的各对立频次；(3)根据音节的各对立频次的数量大小关系，计算出每一个音节的三项对立、两项对立和单项对立分别承载的负担量。

另外，封闭语料库中单音节的总语素数用符号 N 表示，代表整个语料库的负担总量。

(一) 三项对立音位负担量计算

音节三项对立类型的负担量通过音节三项对立频次数量来计算。因此，第 i 个音节的三项对立负担量 FFL_TR(i)，可以通过公式(1)得到：

$$FFL_TR_{(i)} = T_{(i)} * \frac{SYD_{(i)}}{N - T_{(i)}} \quad (1)$$

公式(1)中，SYD 代表三项对立频次。

(二) 两项对立音位负担量计算

音节两项对立音位负担量为三种两项对立方式的负担量之和。因此，第 i 个音节的两项对立负担量 FFL_DB(i)，可以通过公式(2)得到：

$$FFL_DB_{(i)} = T_{(i)} * \frac{SY_{(i)} + SD_{(i)} + YD_{(i)}}{N - T_{(i)}} \quad (2)$$

公式(2)中，SY 代表两项声/韵对立频次，SD 代表两项声/调对立频次，YD 代表两项韵/调对立频次。

(三) 单项对立音位负担量计算

音节单项对立类型的负担量为三种单项对立方式负担量的总和。因此，第 i 个音节单项对立类型负担量 FFL_SG(i)，可以通过公式(3)得到：

$$FFL_SG_{(i)} = T_{(i)} * \frac{S_{(i)} + Y_{(i)} + D_{(i)}}{N - T_{(i)}} \quad (3)$$

公式(3)中，S 代表单项声母对立频次，Y 代表单项韵母对立频次，D 代表单项声调对立频次。

通过上面对立类型负担量的计算公式，可以得到不同音节对立类型上的负担量以及不同音节在对立类型上的差异。以“(pa55)”音节为例，对应 3 个语素“巴、疤、八”，所以“(pa55)”的总负担量为 3，按照对立的频次比例关系分配在除“无对立”的其他 7 种对立条件下。如表 3 所示。

表 3　北京话对立类型负担量计算表

序号	音节	三项对立	两项对立			单项对立			同音字	负担量
		声/韵/调	声/韵	声/调	韵/调	声母	韵母	声调		
1	pa55	2.122	0.681	0.042	0.098	0.017	0.033	0.007	巴、疤、八	3
2	pa35	0.704	0.229	0.014	0.04	0.006	0.003	0.003	拔	1
3	pa214	0.774	0.16	0.017	0.034	0.002	0.009	0.003	把(握)	1
…	…	…	…	…	…	…	…	…	…	…
1036	iuŋ51	0.572	0.303	0.003	0.079	0	0.041	0.002	用	1
平均值		1.941	0.694	0.073	0.109	0.027	0.045	0.008		
最大值		10.568	5.233	0.96	1.282	0.459	0.615	0.157		
最小值		0.56	0.139	0	0.008	0	0.0003	0		
合计		2011.28	718.49	75.68	112.61	27.45	46.06	8.37		3000

2.3 音位负担量计算

语言系统中，每个音节的地位是靠声母、韵母和声调 3 个对立单元产生的 8 种对立关系来体现。音节同音字越多，该音节承载的音位负担量越大，在语言系统中的地位越高，所分配到声母、韵母和声调的音位负担量就越大。

(一) 声母负担量计算

在声/韵/调三项对立分布中，声母承担总负担量的 1/3，在声/韵和声/调两项对立分布中，声母均承担总负担量的 1/2，而在声母单项对立中，声母承担所有负担量。因此，可通过公式(4)对声母负担量 FFL_S(i)来进行计算：

$$FFL_S_{(i)} = T_{(i)} * \frac{SYD_{(i)}/3 + SY_{(i)}/2 + SD_{(i)}/2 + S_{(i)}}{N - T_{(i)}} \tag{4}$$

然后把相同声母的负担量相加，得到某个声母的负担量 FFL_S，P 为某个声母出现的音节个数。见公式(5)：

$$FFL_S = \sum_{i=1}^{P} \left(FFL_S_{(i)} \right) \tag{5}$$

某个声母负担量 FFL_S 除以 P 得到其平均负担量 FFL_S_AV，见公式(6)：

$$FFL_S_AV = \frac{\sum_{i=1}^{P} \left(FFL_S_{(i)} \right)}{P} \tag{6}$$

其他声母也用同样的方法进行计算，得到每一个声母的负担量，所有声母负担量的总和即为声母总负担量。

(二) 韵母负担量计算

在声/韵/调三项对立分布中，韵母承担总负担量的 1/3，在声/韵和韵/调两项对立分布中，韵母均承担总负担量的 1/2，而在韵母单项对立中，韵母承担所有负担量。因此，在对韵母负担量 FFL_Y(i)计算，可通过公式(7)来进行计算：

$$FFL_Y_{(i)} = T_{(i)} * \frac{SYD_{(i)}/3 + SY_{(i)}/2 + YD_{(i)}/2 + Y_{(i)}}{N - T_{(i)}} \tag{7}$$

然后把相同韵母的负担量相加，得到某个韵母的负担量 FFL_Y，P 为某个韵母出现的音节个数，见公式(8)：

$$FFL_Y = \sum_{i=1}^{P} \left(FFL_Y_{(i)} \right) \tag{8}$$

某个韵母负担量 FFL_Y 除以 P 得到其平均负担量 FFL_Y_AV，见公式(9)：

$$FFL_Y_AV = \frac{\sum_{i=1}^{P} \left(FFL_Y_{(i)} \right)}{P} \tag{9}$$

其他韵母也用同样的方法进行计算，得到每一个韵母的负担量，所有韵母负担量的总和即为韵母总负担量。

(三) 声调负担量计算

在声/韵/调三项对立分布中，声调承担总负担量的 1/3，在声/调和韵/调两项对立中，声

调均承担总负担量的 1/2，而在声调单项对立中，声调承担所有负担量。因此，对声调负担量 FFL_D(i)计算，可通过公式(10)来进行计算：

$$FFL_D_{(i)} = T_{(i)} * \frac{SYD_{(i)}/3 + SD_{(i)}/2 + YD_{(i)}/2 + D_{(i)}}{N - T_{(i)}} \tag{10}$$

然后把相同声调的负担量相加，得到某个声调的负担量 FFL_D，P 为某个声调出现的音节个数，见公式(11)：

$$FFL_D = \sum_{i=1}^{P} \left(FFL_D_{(i)} \right) \tag{11}$$

声调负担量 FFL_D 除以 P 得到某个声调的平均负担量 FFL_D_AV，见公式(12)：

$$FFL_D_AV = \frac{\sum_{i=1}^{P} \left(FFL_D_{(i)} \right)}{P} \tag{12}$$

其他声调也用同样的方法进行计算，得到每一个声调的负担量，所有声调负担量的总和即为声调总负担量。

叁　汉语方言音位结构负担计算

根据上一节音位负担的算法，我们对汉语 20 个方言的音位结构负担进行了计算。本文利用该结果来说明音位负担的语言学意义。步骤为：(1)对音节的对立频次进行统计计算，分别得到每个音节在不同对立类型中的各自对立频次；(2)对对立类型进行负担量计算，计算每种对立类型下所包含的对立方式所承担的负担量，为了表述方便将负担量转换成了百分比，下同。见图 2 和表 3；(3)计算出音位即声母、韵母、声调承担的负担量，如图 3 所示；(4)用每一个方言的声母负担量除以声母数量，韵母和声调也用同样的方法，得到声母、韵母和声调的平均负担量，如图 4 所示。

3.1 汉语方言音节对立类型频次

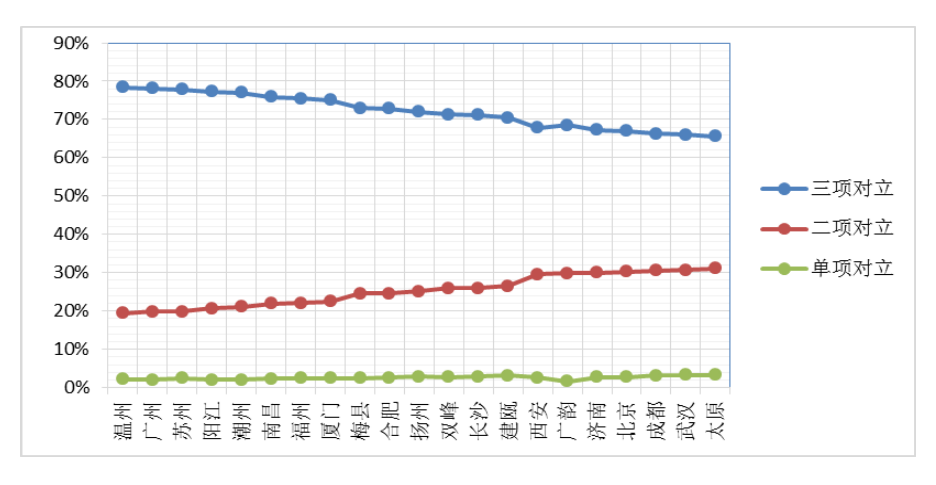


图 2　不同对立方式的频度百分比关系图

从图 2 可以看出：(1)虽然单项对立是语言区别意义的根本，但是单项对立数值很小，汉语方言更多的是靠三项对立和两项对立来完成语意的区别；(2)对立方式负担量有相对稳定的范围；(3)两项对立和三项对立为反比关系。

表 4 不同对立方式的频度百分比（%）

方言点	三项对立	二项对立	单项对立	方言点	三项对立	二项对立	单项对立
温州	78.38	19.49	2.13	双峰	71.33	25.96	2.71
广州	78.09	19.85	2.06	长沙	71.14	25.93	2.93
苏州	77.78	19.79	2.43	建瓯	70.43	26.48	3.09
阳江	77.23	20.66	2.11	西安	67.88	29.57	2.55
潮州	76.98	21.05	1.97	广韵	68.57	29.82	1.61
南昌	75.84	21.88	2.28	济南	67.32	29.94	2.74
福州	75.51	22.09	2.41	北京	67.04	30.23	2.73
厦门	75.06	22.44	2.50	成都	66.23	30.60	3.18
梅县	72.92	24.57	2.50	武汉	65.99	30.73	3.28
合肥	72.82	24.58	2.59	太原	65.63	31.04	3.33
扬州	72. 02	25. 15	2. 83				

从表 4 的数据可以得出：(1)三项对立中温州话最大，为 78.38%，太原最小，为 65.63%，其分布范围有 12.75 个百分点，平均为 72%；(2)两项对立中太原最大，为 31.04%，温州最小，为 19.49%，其分布范围有 11.55 个百分点，平均为 25.27%；(3)单项对立中太原最大，为 3.33%，广韵最小，为 1.61%，其分布范围只有 1.72 个百分点，平均为 2.47%。

在音位学理论上，是否能确定一个音位通常要看是否有最小对立对，即音位学的对立原则，然而从汉语方言的对立类型看，最小对立功能很弱。因此用最小对立来描写汉语方言实际上不能体现出汉语方言的音位功能。音位学中第二项原则是互补原则，从音位对立类型的频度可以看出，两项对立和三项对立包含了互补现象，但大部分并不是由互补造成的。在汉语中 ge、ke、he 与 ji、qi、xi 实际上也没有利用互补原则处理。利用音位对立类型的频度不仅可以描写音位的对立类型和分布，也可以描写音位功能的类型，为研究语言的音位结构、信息功能、语言接触和融合以及历史演化开辟了新的思路和新的研究方法。

3.2 汉语方言音位总负担量

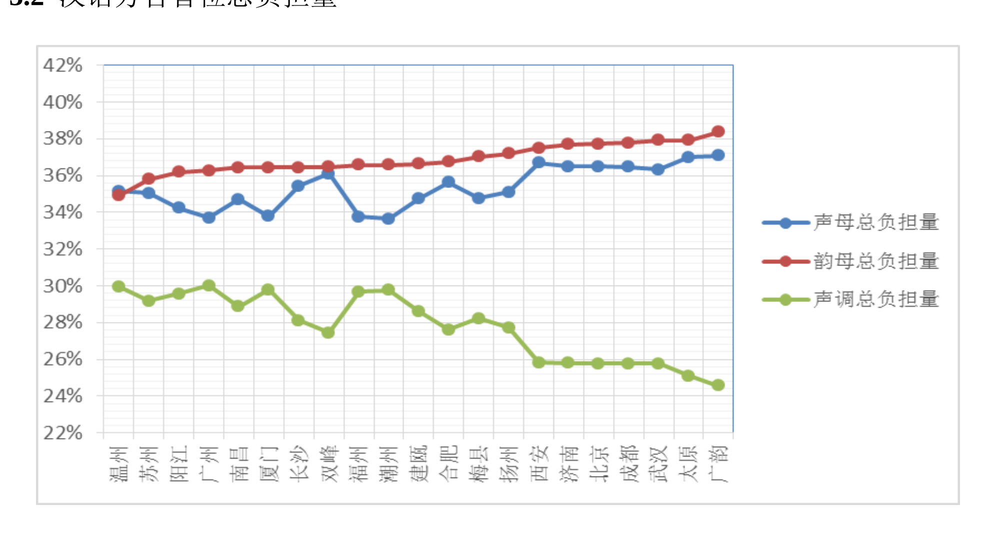


图 3　汉语方言声韵调总负担关系图

从图 3 可以看出：(1)汉语方言声母和韵母的负担量较大，其中韵母略大于声母的负担量，声调较小；(2)汉语方言声母、韵母和声调的总负担量各自有一定的范围；(3)声母和韵母的总负担量与声调的总负担量为反比关系，即声母和韵母的总负担量越大，声调的总负担

量就越小。

表 5 音位总负担量百分比（%）

方言点	声母负担量	韵母负担量	声调负担量	方言点	声母负担量	韵母负担量	声调负担量
温州	35.15	34.90	29.95	合肥	35.64	36.75	27.61
苏州	35.03	35.81	29.16	梅县	34.75	37.02	28.23
阳江	34.24	36.18	29.57	扬州	35.10	37.19	27.71
广州	33.71	36.28	30.01	西安	36.70	37.49	25.81
南昌	34.71	36.43	28.86	济南	36.50	37.70	25.80
厦门	33.79	36.43	29.78	北京	36.50	37.74	25.77
长沙	35.43	36.45	28.12	成都	36.45	37.79	25.75
双峰	36.11	36.45	27.44	武汉	36.32	37.91	25.77
福州	33.75	36.58	29.67	太原	36.99	37.92	25.09
潮州	33.65	36.59	29.76	广韵	37.07	38.37	24.56
建瓯	34.75	36.63	28.62				

从表 5 可以看出：(1)汉语方言声母的总负担量中，广韵最大，为 37.07%，潮州最小，为 33.65%。其分布的范围有 3.42 个百分点，平均为 35.35%；(2)汉语方言韵母的总负担量中，广韵最大，为 38.37%，温州最小，为 34.9%。其分布的范围有 3.47 个百分点，平均为 36.89%；(3)汉语方言声调的总负担量中，广州最大，为 30.01%，广韵最小，为 24.56%。其分布的范围有 5.45 个百分点，平均为 27.76%。

从汉语方言声母、韵母和声调的总负担量来看，它们在各个方言中存在差异。首先，声母、韵母和声调的总负担量是不同的，有的大，有的小，反映出了音位结构和功能的差异。其次，声母和韵母的性质十分相同，总体上看，韵母的总负担量略大于声母的总负担量，而声调的总负担量要比声母和韵母的总负担量小很多。第三，从数据上看，声母和韵母的总负担量大，声调的总负担量就小。这些数据体现出，汉语方言声、韵、调三者之间是一种动态的和有机的关系，同时也体现出了汉语方言音位结构的变化。由于选用汉语方言的语素相同，要表达的信息量也相同，但在长期的演化过程中，由于各种因素的影响，每个方言的音位结构系统都会受到影响，因此，音位负担量的描写是一种对音位动态系统和功能精确的数字化描写。

3.3 汉语方言音位平均负担量

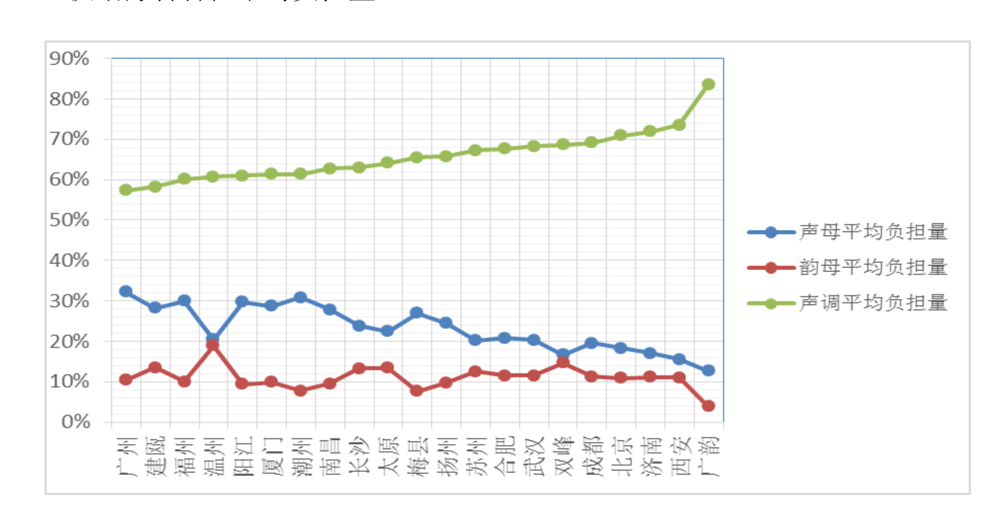


图 4 音位系统的平均音位负担量关系图

从图 4 可以看出：(1)汉语方言声母和韵母的平均负担量较小，其中声母略大于韵母的负担量，声调较大；(2)汉语方言声母、韵母和声调的平均负担量各自有一定的范围；(3)声调的平均负担量与声母和韵母的平均负担量为反比关系，即声调的平均负担量越大，声母和韵母的平均负担量越就越小。

表 6　音位平均负担量百分比（%）

方言点	声母平均	韵母平均	声调平均	方言点	声母平均	韵母平均	声调平均
广州	32.23	10.41	57.37	扬州	24.48	9.80	65.72
建瓯	28.26	13.54	58.19	苏州	20.19	12.57	67.24
福州	29.92	9.98	60.11	合肥	20.79	11.55	67.66
温州	20.37	18.88	60.75	武汉	20.25	11.48	68.27
阳江	29.72	9.32	60.96	双峰	16.72	14.70	68.59
厦门	28.68	9.92	61.40	成都	19.58	11.27	69.15
潮州	30.86	7.74	61.40	北京	18.24	10.92	70.84
南昌	27.77	9.55	62.68	济南	16.97	11.07	71.96
长沙	23.80	13.23	62.97	西安	15.49	10.96	73.55
太原	22.49	13.45	64.06	广韵	12.61	3.87	83.53
梅县	26.89	7.58	65.53				

从表 6 可以看出：(1)汉语方言声母的平均负担量中，广州最大，为 33.23%，广韵最小，为 12.61%。其分布的范围有 20.62 个百分点，平均为 22.92%；(2)汉语方言韵母的平均负担量中，温州最大，为 18.88%，广韵最小，为 3.87%。其分布的范围有 15.01 个百分点，平均为 11.38%；(3)汉语方言声调的平均负担量中，广韵最大，为 83.53%，广州最小，为 57.37%。其分布的范围有 21.16 个百分点，平均为 70.45%。

从汉语方言声母、韵母和声调的平均负担量性质看，首先是声调的平均负担量数值较大；其次是声母和韵母的平均负担量较小，其中声母的平均负担量略大于韵母的平均负担量；第三是声调的平均负担量与声母和韵母的平均负担量成反比关系。从整体上看，平均负担量和总负担量体现为相反的性质，即总负担量高的方言，其平均负担量要小，而总负担量小的语言平均负担量反而要大。例如，广州话的声调数量多，总负担量较大，但平均下来，广州话声调的平均负担量反而没有北京话的大，仔细分析可以发现，广州话的声调和声韵母有大量互补分布，虽然同是 3000 语素，实际声调的功能要弱。

平均负担量和总负担量不同，总负担量和音位的数量有关，平均负担量不受音位数量的影响。在一个音位系统中，平均负担量表示音位系统中每个音位的负担，音位的平均负担量大说明要区别的语素多，也会造成同音词多。本文的重点只是解释音位结构负担的基本定义和算法，没有涉及到汉语方言音位负担量的细节内容。实际上，一个音位系统中，每个音位的总负担量和平均负担量都是不同的，基本上体现为一种近似指数的分布（孔江平 2013）。因此，在共时音位系统和历时音变中，一个音位的负担量大可以形成语音规则，负担量小则不宜形成语音规则，而且很可能容易丢失[1]。在此举一个例子只是说明音位结构负担量能解释大量语言和方言在共时系统和历时演变的本质。

[1] 汉语方言负担量研究将另文发表。

肆 结束语

根据本文提出的这种音位负担的算法，我们计算了汉、藏、缅声韵调的音位负担，还对汉语方言和藏语方言进行了计算，揭示了许多和以往不同的语言音位结构、分布和功能的规律。我们相信通过负担量计算比较，可以探讨某一音位组合或某一音位在具体语言系统中的地位问题。对汉藏语系语言来说，就是声母、韵母和声调的组合方式和他们各自的组合能力强弱问题。虽然，我们可以通过统计每一个音位出现的次数，来看在语言系统中具体音位的使用地位，但并不能真正的描述和解释每一声韵调的音韵地位和相互组合关系。

总而言之，本研究一方面是以单音节的声韵调音位系统为计算对象，计算音位单元的对立频次和负担量，这种方法如何应用到词汇、短语和句子等层面的语言负担量计算，还需要深入探讨，特别是自然语流中存在的连续音变、协同发音、前后音位相互影响等特殊语音现象。目前对语言负担量的贡献还无法量化。另一方面，这种计算方法，如何进一步扩展到语言类型学、历史语言学、语言习得和语音识别中的应用以及语料设计和后期的语言模型优化等方面，还需进一步深入研究。

参考文献

参考文献

北京大学中国语言文学系语言学教研室编，王福堂修订 2003 《汉语方音字汇》，(北京)语文出版社.

布龙菲尔德，袁家骅、赵世开、甘世福译，钱晋华校. 1980 《语言论》,(北京)商务印书馆.

布洛赫，B.、G. L. 特雷杰,赵世开译 1965 《语言分析纲要》,(北京)商务印书馆.

孔江平 2013 音位负担计量研究——以藏缅语为例 [A]. 石锋，彭刚[主编].《大江东去：王士元教授八十岁贺寿文集》,(香港)香港城市大学出版社 .215—220.

萨丕尔，陆卓元译，陆志韦校. 1985 《语言论》,(北京)商务印书馆.

赵元任，叶蜚声译 1985 音位标音法的多能性 [A]. 赵元任.《赵元任语言学论文选》,(北京)中国社会科学出版社.

Vilem Mathesius. 1929. La Structure Phonologique du Lexique du Tch`eque Moderne. Travaux du Cercle Linguistique de Prague, 1:67–84.

Roman Jakobson. 1931. Prinzipien der Historischen Phonologie. Travaux du Cercle Linguistiquede Prague, 4:246–267.

Nikolai Trubetzkoy. 1939. Grundz ̈uge der Phonologie. Travaux du Cercle Linguistique de Prague,7.

Charles F. Hockett. 1955. A Manual of Phonology (= International Journal of American Linguis-tics 21. Indiana University Press.

Charles F. Hockett. 1967. The Quantification of Functional Load. Word, 23:320–339.

Greenberg, H. H. 1959 A method of measuring functional yield as applied to tone in African languages. Georgetown University, Monograph Series on Languages and Linguistics, 12.7-16

Hoenigswald, H. M. 1965 Language change and linguistic reconstruction. Chicago: University of Chicago Press,

William S-Y. Wang. 1967. The Measurement of Functional Load. Phonetica, 16:36–54.

King, R. D. 1965 Functional load: its measure and its role in sound change. University of Wisconsin PhD dissertation.

Robert King. 1967a. Functional Load and Sound Change. Language, 43.

Robert King. 1967b. A Measure for Functional Load. Studia Linguistica, 21:1-14.

Shannon, C.E. and W. Weaver 1949 The mathematical theory of communication. Urbana: University of Illinois Press.19(7): 1.

Claude E. Shannon. 1951. Prediction and Entropy of Printed English. Bell Systems Technical Journal, 30:50–64.

Henry Kucera. 1963. Entropy, Redundancy and Functional Load. In American Contributions to the Fifth International Conference of Slavists, pages 191–219, Sofia, Bulgaria. The Hague: Mouton.

Dinoj Surendran and Partha Niyogi. 2003. Measuring the Usefulness (Functional Load) of Phonological Contrasts. Technical Report TR-2003-12., Department of Computer Science, University of Chicago.

Dinoj Surendran and Gina-Anne Levow. 2004. The Functional Load of Tone in Mandarin is as High as that of Vowels. In Proceedings of the International Conference on Speech Prosody 2004, 99–102

Abstract

According to the phoneme theory in the structural linguistics and based on the databases of different languages and dialects in China, the phoneme data analysis and series of theoretical studies have been done and then the calculating methods on functional load of phoneme structure have been developed. The main calculating methods are: 1) the calculating on the frequency of syllable contrast; 2) the calculating on the contrast type of phoneme load; 3) the calculating on the load of phonemic system; 4) the calculating on the average phoneme load. By these methods, the contrast type of phoneme load, the load of phonemic system and the average phoneme load of the 3000 single syllable words in 20 Chinese dialects are calculated and studied. Based on these methods and results, the basic theory of phoneme load are proposed and discussed. These methods can be used in the study of phoneme loads of different languages, the sound change in historical linguistics, language evolution and the application of speech engineering.

Keywords:

Phoneme, functional load, phoneme load, phoneme structure